ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Spring 2013
Instructor: Victor Aguirregabiria

SOLUTION TO FINAL EXAM
Friday, April 12, 2013. From 9:00-12:00 (3 hours)

INSTRUCTIONS: The exam consists of two sets of questions. Try to answer all the questions. No study aids, including calculators, are allowed.
TOTAL MARKS = 100

PROBLEM 1 (60 points). Consider the Binary Choice Logit model, $Y = 1\{X'\ \beta - \varepsilon > 0\}$, where $1\{.\}$ is the indicator function, and ε is independent of X with Logistic distribution with mean zero and unit variance, such that $\Pr(\varepsilon \leq c) = \Lambda(c) \equiv \exp\{c\}/[1 + \exp\{c\}]$. Let $\{y_i, x_i : i = 1, 2, ..., n\}$ be a random sample of the variables (Y, X).

(a) [5 points] Write the expression for the log-likelihood function for this model and data.

ANSWER. Given that we have a random sample, $\Pr(y_1,\, y_2,\, ..., y_n \mid x_1,\, x_2,\, ...,\, x_n) = \prod_{i=1}^{n} \Pr(y_i, \mid x_i)$. The model implies that, $\Pr(Y = 1 | \ X = x_i) = \Pr(\varepsilon_i < x_i'\beta) = \Lambda(x_i'\beta)$, and $\Pr(Y = 0 \mid X = x_i) = \Pr(\varepsilon_i \geq x_i'\beta) = 1 - \Lambda(x_i'\beta)$. Therefore, the likelihood function is:

$$L(\beta) = \prod_{i:y_i=1} \Lambda(x_i'\beta) \prod_{i:y_i=0} [1 - \Lambda(x_i'\beta)]$$

And the log-likelihood function is:

$$\ell(\beta) = \sum_{i=1}^{n} y_i \ \ln \Lambda(x_i'\beta) + (1 - y_i) \ \ln[1 - \Lambda(x_i'\beta)]$$

(b) [5 points] Derive the expression for the likelihood-equations that implicitly define the Maximum Likelihood estimator, and show that this system of equations is: $\sum_{i=1}^{n} x_i[y_i - \Lambda\,(x_i'\beta)] = 0$.

ANSWER. The likelihood equation is the system of first order conditions of optimality for the MLE: $\partial\ell(\beta)/\partial\beta = 0$. Using the previous expression for the log-likelihood function, we have:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} y_i \ \frac{\Lambda'(x_i'\beta)}{\Lambda(x_i'\beta)} x_i - (1 - y_i) \ \frac{\Lambda'(x_i'\beta)}{1 - \Lambda(x_i'\beta)} x_i = 0$$

where $\Lambda'(.)$ is the derivative of the CDF $\Lambda(.)$, i.e., the density function of the logistic distribution. Operating in this equation we get:

$$\sum_{i=1}^{n} \Lambda'(x_i'\beta) \ x_i \left[\frac{y_i}{\Lambda(x_i'\beta)} - \frac{1 - y_i}{1 - \Lambda(x_i'\beta)} \right] = 0$$

and

$$\sum_{i=1}^{n} \frac{\Lambda'(x_i'\beta)}{\Lambda(x_i'\beta)\ [1-\Lambda(x_i'\beta)]}\ x_i\,[y_i - \Lambda(x_i'\beta)] = 0$$

Finally, by definition of the function $\Lambda(c) \equiv \exp\{c\}/[1+\exp\{c\}]$, we have that $\Lambda'(c) = \exp\{c\}/[1+\exp\{c\}]^2$ such that $\Lambda'(c) = \Lambda(c)\ [1-\Lambda(c)]$. Therefore, the system of likelihood equations is:

$$\sum_{i=1}^{n} x_i\ \ [y_i - \Lambda(x_i'\beta)] = 0$$

(c) [5 points] Explain why this MLE can be interpreted as a Methods of Moments estimator. What are the moment conditions?

ANSWER. The MLE is the value of β that solves the system of equations $n^{-1}\sum_{i=1}^{n} x_i\ [y_i - \Lambda(x_i'\beta)] = 0$. The vector $n^{-1}\sum_{i=1}^{n} x_i\ [y_i - \Lambda(x_i'\beta)]$ is the sample mean of $x_i\ [y_i - \Lambda(x_i'\beta)]$. Therefore, these conditions are the sample counterpart of the population moment conditions $E(\ X\ [Y - \Lambda(X'\ \beta)]\ ) = 0$. The estimator that solves the sample counterpart of a set of population moment conditions is a MM estimator. Therefore, the MLE can be interpreted as a MM estimator.

(d) [10 points] Obtain the expression for Hessian matrix of the log-likelihood function. Show that the log-likelihood function is globally concave (i.e., for any value of β).

ANSWER. By definition, the Hessian matrix is $H \equiv \partial^2\ell(\beta)/\partial\beta\ \partial\beta'$ such that

$$H = \frac{\partial\ell(\beta)/\partial\beta}{\partial\beta'} = \frac{\partial}{\partial\beta'}\left(\sum_{i=1}^{n} x_i\,[y_i - \Lambda(x_i'\beta)]\right) = -\sum_{i=1}^{n} x_i\ x_i'\ \Lambda'(x_i'\beta) = -\sum_{i=1}^{n}(x_i^*\ x_i^{*\prime})$$

where $x_i^* \equiv x_i\ \Lambda'(x_i'\beta)$. Provided that there is not perfect collinearity between the vector of explanatory variables in X, it is clear that $\sum_{i=1}^{n}(x_i^*\ x_i^{*\prime})$ is positive definite matrix, such as the Hessian $H = -\sum_{i=1}^{n}(x_i^*\ x_i^{*\prime})$ is negative definite. This implies that the log-likelihood function is globally concave in β.

(e) [10 points] Describe the BHHH iterative algorithm (or alternatively, the Newton-Raphson algorithm) for the computation of the MLE. Write the formula describing an iteration in this algorithm. Explain how to implement the algorithm.

ANSWER. The BHHH is an iterative algorithm to compute the maximum of the log-likelihood function. It is based on the following formula: at iteration $k \geq 1$, we obtain the vector $\widehat{\beta}_k$ as:

$$\widehat{\beta}_k = \widehat{\beta}_{k-1} + \left[\sum_{i=1}^{n} s_i(\widehat{\beta}_{k-1})\ s_i(\widehat{\beta}_{k-1})'\right]^{-1}\left[\sum_{i=1}^{n} s_i(\widehat{\beta}_{k-1})\right]$$

where $s_i(\beta)$ is the vector of scores, i.e., the contribution of observation i to the likelihood equations. In this model we have that $s_i(\beta) = x_i\ [y_i - \Lambda(x_i'\beta)]$, and therefore,

$$\widehat{\beta}_k = \widehat{\beta}_{k-1} + \left[\sum_{i=1}^{n} x_i\ x_i'\ \left(y_i - \Lambda(x_i'\widehat{\beta}_{k-1})\right)^2\right]^{-1}\left[\sum_{i=1}^{n} x_i\left(y_i - \Lambda(x_i'\widehat{\beta}_{k-1})\right)\right]$$

The algorithm is implemented in the following way. We start with an arbitrary initial value $\widehat{\beta}_0$. Then, we apply recursively the previous formula that describes at iteration, for $k = 1, 2, ...$ At every iteration k, we calculate the Euclidean norm $||\widehat{\beta}_k - \widehat{\beta}_{k-1}||$. If this distance is lower than a pre-specified small constant (e.g., say 10^{-6}), we stop the algorithm and use $\widehat{\beta}_k$ as the MLE. Otherwise we continue with iteration $k + 1$.

(f) [5 points] Let $\widehat{\beta}$ be the ML estimate of β. Based on this estimate, describe a test of the null hypothesis "ε is independent of X with Logistic distribution Λ."

ANSWER. According to the model, $E(Y \mid X = x_i) = \Lambda(x_i'\beta)$. Therefore, if the specification of the distribution Λ is correct we have that $y_i = \Lambda(x_i'\beta) + u_i$, where $E(u_i|x_i) = 0$. We can construct a test for this prediction of the model. The null hypothesis is that $E(u_i|x_i) = 0$. Given $\widehat{\beta}$, we can construct the residuals: $\widehat{u}_i = y_i - \Lambda(x_i'\widehat{\beta})$. Note that by construction of $\widehat{\beta}$ (by the likelihood equations) we have that $\sum_{i=1}^n x_i\ \widehat{u}_i = 0$. Therefore, we cannot test the null hypothesis $E(x_i\ u_i) = 0$ because by construction the residuals are not correlated with X. This is exactly the same result as in a linear regression model. However, we can test for $E(h(x_i)\ u_i) = 0$ where $h(x_i)$ is a vector of nonlinear functions of x_i, i.e., power functions, cross-products, etc. We run an OLS regression of $\widehat{u}_i$ on the vector $h(x_i)$. Let $\widehat{\chi}$ be $n\ R^2$, where R^2 is the R-square coefficient in this regression. Under the null hypothesis $\widehat{\chi}$ is asymptotically distributed as a Chi-square random variable with degrees of freedom equal to the dimension of the vector $h(x_i)$. We reject the null hypothesis with a confidence level of $\alpha\%$ if $\widehat{\chi}$ is greater than the critical value χ_α, where χ_α is $1 - \alpha$ quantile of Chi-square(q) distribution.

(g) [10 points] Suppose that you reject the null hypothesis in the previous test. Describe how to use Klein-Spady semiparametric method to obtain a consistent estimator of (β, F), where $F(.)$ is a nonparametrically specified distribution of ε.

ANSWER. If the vector x_i contains at least one continuous explanatory variable, we can apply Klein-Spady estimator. We can implement this estimator using the following iterative procedure. At every iteration $t \geq 1$, we obtain the estimates $\{\widehat{F}_t, \widehat{\beta}_t\}$ using the following formulas. Given $\widehat{\beta}_{t-1}$, we update our estimate of the distribution F at any possible value z using the following kernel estimator:

$$\widehat{F}_t(z) = \frac{\sum_{i=1}^n y_i\ K\left(\frac{z - \widehat{z}_i^{t-1}}{b}\right)}{\sum_{i=1}^n K\left(\frac{z - \widehat{z}_i^{t-1}}{b}\right)}$$

where, for every observation i, $\widehat{z}_i^{t-1}$ is the index $x_i'\widehat{\beta}_{t-1}$, $K(.)$ is a kernel function and b is a pre-specified bandwidth parameter. Then, given $\widehat{F}_t(.)$, we update our estimate of β by obtaining $\widehat{\beta}_t$ as the value of β that maximizes the log-likelihood function where we use $\widehat{F}_t(.)$ as the distribution of ε. At every iteration t, we calculate the Euclidean norm $||\widehat{\beta}_t - \widehat{\beta}_{t-1}||$. If this distance is lower than a pre-specified small constant (e.g., say 10^{-6}), we stop the algorithm and use $\{\widehat{F}_t, \widehat{\beta}_t\}$ as our semiparametric MLE. Otherwise we continue with iteration $t + 1$. Note that, by construction, convergence in $\{\widehat{\beta}_t\}$ implies convergence in $\{\widehat{F}_t\}$ and viceversa.

(h) [10 points] Suppose that the assumption "ε is independent of X with Logistic distribution Λ" cannot be rejected. However, the researcher considers that one of the regressors in X, say

W, is potentially endogenous such that W and ε are not independently distributed. Explain how to construct a test of the null hypothesis "W and ε are not independently distributed". [Note: Normality of ε is not necessary to implement the test.]

ANSWER. Let $X = (\widetilde{X}, W)$, i.e., $\widetilde{X}$ represents the exogenous regressors in X. Suppose that there is an observable variable (or vector of variables) Z such that: (1) Z that does not enter in $\widetilde{X}$ and it is not perfectly collinear with $\widetilde{X}$; (2) we are willing to assume that Z is independent of ε; and (3) Z is correlated with W even after controlling for the correlation between W and $\widetilde{X}$. Under these conditions, we can test for the endogeneity of W in the binary choice model. The null hypothesis is that W and ε are independent. The procedure to implement the test is the following. First, we ran an OLS regression for W on $\widetilde{X}$ and Z. Let $\widehat{w}_i$ be the fitted values from this regression. Second, we estimate by ML a logit model for y_i with explanatory variables $\widetilde{x}_i$, w_i, and $\widehat{w}_i$. Under the null hypothesis, the coefficient associated with $\widehat{w}_i$ should be zero. This null hypothesis can be tested simply using a t-ratio test for that coefficient. Note that we have not made any assumption about the distribution of the error term in the regression model for W. The form of that distribution is relevant for the consistent estimation of the parameters in the second step, but it does not affect the validity of this test (Rivers and Vuong, 1988).

PROBLEM 2 (40 points). Consider the following Panel Data system of simultaneous equations:

$$
\begin{aligned}
y_{it} &= \beta_1 \, x_{it} + \beta_2 \, y_{it-1} + \alpha_i + \lambda_t + u_{it} \\
x_{it} &= \gamma_1 \, y_{it} + \gamma_2 \, x_{it-1} + \eta_i + \delta_t + v_{it}
\end{aligned}
$$

where u_{it} and v_{it} are not serially correlated, but they can be correlated with each other. The individual effects α_i and η_i can be also correlated with each other. The panel dataset contains a large number N of individuals but a small number of time periods, say $T = 4$.

(a) [15 points] Show whether each of the following estimators is consistent as $N \to \infty$ and fixed T. (1) OLS in levels; (2) Within-Groups estimator; (3) OLS in first differences; (4) IV estimation of the equation for y_{it} in levels using as instruments $\{y_{it-1}, x_{it-1}\}$; (5) IV estimation of the equation for y_{it} in levels using as instruments $\{x_{it-2}, x_{it-3}\}$; (6) IV estimation of the equation y_{it} in first differences using as instruments $\{x_{it-1}, y_{it-2}\}$; and (7) IV estimation of the equation y_{it} in first differences using as instruments $\{x_{it-2}, y_{it-2}\}$.

ANSWER.

(1) OLS in levels. Inconsistent because $E(\{\alpha_i + u_{it}\} \; y_{it-1}) \neq 0$ and $E(\{\alpha_i + u_{it}\} \; x_{it}) \neq 0$.

(2) Within-Groups estimator. Inconsistent. In dynamic PD models with small T, the WG estimator is inconsistent.

(3) OLS in first differences. Inconsistent. In dynamic PD models with small T, the OLS-FD estimator is inconsistent.

(4) IV estimation of the equation for y_{it} in levels using as instruments $\{y_{it-1}, x_{it-1}\}$. Inconsistent because $E(\{\alpha_i + u_{it}\}\ y_{it-1}) \neq 0$ and $E(\{\alpha_i + u_{it}\}\ x_{it-1}) \neq 0$.

(5) IV estimation of the equation for y_{it} in levels using as instruments $\{x_{it-2}, x_{it-3}\}$. Inconsistent because $E(\{\alpha_i + u_{it}\}\ x_{it-2}) \neq 0$ and $E(\{\alpha_i + u_{it}\}\ x_{it-3}) \neq 0$ unless $\gamma_1 = 0$ and η_i and α_i are not correlated.

(6) IV estimation of the equation y_{it} in first differences using as instruments $\{x_{it-1}, y_{it-2}\}$. Inconsistent because $E(\Delta u_{it}\ x_{it-1}) \neq 0$ because $E(u_{it-1}\ x_{it-1}) \neq 0$ unless $\gamma_1 = 0$ and u_{it-1} and v_{it-1} are not correlated.

(7) IV estimation of the equation y_{it} in first differences using as instruments $\{x_{it-2}, y_{it-2}\}$. Consistent because $E(\Delta u_{it}\ x_{it-2}) \neq 0$ and $E(\Delta u_{it}\ y_{it-2}) \neq 0$ as long as u_{it} and v_{it} are not serially correlated. We also need that $\beta_2 \neq 0$ and $\gamma_2 \neq 0$ for these instruments to have predictive power.

(b) [15 points] Consider the estimation of β_1 and β_2 in the equation for y_{it} using Arellano-Bond GMM estimator. Write the expressions of all the 'Arellano-Bond' moment conditions in this model and dataset with $T = 4$ time periods.

ANSWER. With $T = 4$ we have two equations (two time periods) in the model in first differences.

$$\begin{aligned}
\Delta y_{i3} &= \beta_1\ \Delta x_{i3} + \beta_2\ \Delta y_{i2} + \Delta\lambda_3 + \Delta u_{i3} \\
\Delta y_{i4} &= \beta_1\ \Delta x_{i4} + \beta_2\ \Delta y_{i3} + \Delta\lambda_4 + \Delta u_{i4}
\end{aligned}$$

For the first equation we have the following moment conditions: $E(\Delta u_{i3}) = 0$, $E(y_{i1}\ \Delta u_{i3}) = 0$, and $E(x_{i1}\ \Delta u_{i3}) = 0$. For the second equation we have the following moment conditions: $E(\Delta u_{i4}) = 0$, $E(y_{i1}\ \Delta u_{i4}) = 0$, $E(y_{i2}\ \Delta u_{i4}) = 0$, $E(x_{i1}\ \Delta u_{i4}) = 0$, and $E(x_{i1}\ \Delta u_{i4}) = 0$. In total we have 8 moment conditions to estimate the vector of parameters $\theta = (\beta_1,\ \beta_2,\ \Delta\lambda_3,\ \Delta\lambda_4)'$. Let $m_N(\theta)$ be the vector with the 8 sample moment conditions such that

$$m_N(\theta) = N^{-1} \sum_{i=1}^{N} \begin{bmatrix} (\Delta y_{i3} - \beta_1\ \Delta x_{i3} - \beta_2\ \Delta y_{i2} - \Delta\lambda_3) \\ y_{i1}\ (\Delta y_{i3} - \beta_1\ \Delta x_{i3} - \beta_2\ \Delta y_{i2} - \Delta\lambda_3) \\ x_{i1}\ (\Delta y_{i3} - \beta_1\ \Delta x_{i3} - \beta_2\ \Delta y_{i2} - \Delta\lambda_3) \\ (\Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4) \\ y_{i1}\ (\Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4) \\ x_{i1}\ (\Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4) \\ y_{i2}\ (\Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4) \\ x_{i2}\ (\Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4) \end{bmatrix}$$

We can represent this moment conditions as

$$N^{-1} \sum_{i=1}^{N} Z_i \begin{bmatrix} \Delta y_{i3} - \beta_1\ \Delta x_{i3} - \beta_2\ \Delta y_{i2} - \Delta\lambda_3 \\ \Delta y_{i4} - \beta_1\ \Delta x_{i4} - \beta_2\ \Delta y_{i3} - \Delta\lambda_4 \end{bmatrix}$$

where

$$Z_i = \begin{bmatrix} 1 & 0 \\ y_{i1} & 0 \\ x_{i1} & 0 \\ 0 & 1 \\ 0 & y_{i1} \\ 0 & x_{i1} \\ 0 & y_{i2} \\ 0 & x_{i2} \end{bmatrix}$$

(c) [10 points] Explain how to obtain an Arellano-Bond 1-step-GMM estimator, and a 2-step optimal GMM estimator.

The one-step AB estimator is obtained as the value of θ that minimizes the criterion function $m_N(\theta)'$ $\mathbf{W}^{(1)}$ $m_N(\theta)$, where $\mathbf{W}^{(1)}$ is the matrix $\left[N^{-1}\sum_{i=1}^{N} Z_i \; H \; Z_i'\right]^{-1}$ where H is the matrix $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. The matrix H represents, up to a scale parameter, the variance-covariance matrix of $(\Delta u_{i3},\, \Delta u_{i4})$ under the assumptions of homoscedasticity and no serial correlation in u_{it}. It is the optimal weighting matrix under these assumptions. We can relax the assumption of time-homoscedasticity in the second step. Given the residuals in the first step, say $(\widehat{\Delta u_{i3}},\, \widehat{\Delta u_{i4}})$, we can obtain:

$$\mathbf{W}^{(2)} = \left(N^{-1}\sum_{i=1}^{N} Z_i \; \left[(\widehat{\Delta u_{i3}}, \widehat{\Delta u_{i4}}) \begin{pmatrix} \widehat{\Delta u_{i3}} \\ \widehat{\Delta u_{i4}} \end{pmatrix}\right] Z_i'\right)^{-1}$$

Then, the two-step AB estimator is obtained as the value of θ that minimizes the criterion function $m_N(\theta)'$ $\mathbf{W}^{(2)}$ $m_N(\theta)$. This estimator is asymptotically efficient.